

### **Abstract**

Externalizing behavior in preschool has been found to predict maladjustment later in life. Therefore, it is important to identify children most at risk for continuing externalizing behavior beyond preschool. To date, a number of questionnaires are available for teachers to assist in identifying those children. A frequently overlooked aspect in this screening process is the consideration of different dimensions of externalizing behavior instead of the use of broadband scales. Therefore, a brief, user-friendly teacher questionnaire was adapted to capture different dimensions of externalizing behavior (hyperactivity, opposition, and physical aggression). First, the a priori three-factor structure of this questionnaire was assessed in a large sample of preschoolers ( $N = 3610$ ). Second, factorial invariance of the questionnaire over child gender and child home language was investigated. Results confirmed the three-factor structure of the questionnaire. Configural, metric, and scalar invariance was found for child gender and child home language, which indicates that teachers assigned the same meaning to the three externalizing behavior dimensions across these groups.

Key words: Assessment, Externalizing behavior, Factorial invariance, Screening, Gender, Home language

## A Multidimensional Screening Tool for Preschoolers with Externalizing Behavior: Factor Structure and Factorial Invariance

Externalizing behavior (EB) in early childhood refers to a range of behaviors that are disruptive and/or harmful for others (Goossens, Bokhorst, Bruinsma, & Van Boxtel, 2002). This pattern of behavior is a risk factor for maladjustment in several domains in adolescence and adulthood, such as delinquency, school failure, and mental disorders (Nagin & Tremblay, 1999). Therefore, it is important to identify those children at risk for maladaptive outcomes beyond preschool. A number of teacher questionnaires are available to early childhood professionals (e.g., teachers) to assist in accurately identifying those children (e.g., Loeber & Farrington, 2000; Willoughby, Kupersmidt, & Bryant, 2001). Nevertheless, a frequently overlooked aspect in this screening process is the consideration of different EB-dimensions (e.g., hyperactivity, opposition, physical aggression). While some EB-dimensions may be normative at certain ages, others may be precursors of future maladjustment (Loeber & Farrington, 2000). Screening these different EB-dimensions may be a first step in more accurately identifying those preschoolers most at risk for negative outcomes later in life (Willoughby et al., 2001). The present study sought to test the factor structure of a brief, user-friendly teacher questionnaire targeting hyperactivity, opposition, and physical aggression for preschoolers. Moreover, the present study aimed to investigate the questionnaire's factorial invariance over different groups (boys versus girls, children speaking the majority home language versus all other children).

### **Teacher Screening Instruments for Multidimensional Assessment of Preschoolers' Externalizing Behavior**

For preschoolers, parent ratings and behavior observations are important sources for determining EB (Matthijs & Lochman, 2010). When children enter school, however, teachers become important adult figures who yield unique information concerning child EB in the

school context. Teachers have been shown to provide valid and reliable EB-reports (Konold & Pianta, 2007) and they may play an important role in the detection and timely referral of at risk children (Zwirs et al., 2011).

To date, a number of questionnaires are available for teachers to assess multiple areas of preschoolers' adjustment (e.g., Teacher Report Form, Behavioral Assessment System for Children; see Dever & Kamphaus, 2013, for an overview). However, most of these instruments are extensive in length. In the context of screening, it will be especially difficult for teachers to fill out lengthy questionnaires for several or all preschoolers in their classes (Feil, Walker, & Severson, 1995). A notable exception is the widely used Preschool Behavior Questionnaire (PBQ; Behar & Stringfield, 1974).

However, similar to most teacher screening questionnaires (Willoughby et al., 2001), the PBQ provides a general view on preschoolers' EB. On the one hand, scholars have put forward that EB may be least differentiated in early childhood (Weis, Lovejoy, & Lundahl, 2005). Therefore, broadband externalizing scales may be useful to get a quick view on the overall EB-level. On the other hand, these broadband scales may mask important differences in the presentation and developmental course of different EB-dimensions (Willoughby et al., 2001). To date, only one study aimed at identifying different EB-dimensions within the PBQ and distinguished Physical aggression and Non-aggressive antisocial behavior as subscales of the Externalizing Scale (Spilt, Koomen, Thijs, Stoel, & van der Leij, 2010).

Complementing these authors' work and following research on the developmental course and correlates of different externalizing behavior dimensions (Nagin & Tremblay, 1999), we aimed to distinguish other dimensions of externalizing behavior, i.e., hyperactivity, physical aggression, and opposition. Indeed, research has shown that these three different types of EB in preschool have partly similar (Moreland & Dumas, 2008), but also different correlates later in life. For example, Nagin and Tremblay (2009) showed differential links of

these three behaviors in preschool with juvenile delinquency. Moreover, Harvey, Youngwirth, Thakar, and Errazuriz (2009) found that teacher rating scales on hyperactivity and aggression filled out for 3-year-old preschoolers accurately predicted the diagnosis of ADHD and ODD/CD respectively three years later. Hence, it is important to be able to distinguish these three dimensions at a young age. This distinction may improve predictive accuracy, contribute to developmental research concerning specific risk factors for certain outcomes or diagnoses (Nagin & Tremblay, 1999), and add to more targeted interventions (Tremblay, 2010).

In this study, we developed a questionnaire that was based on the Externalizing scale of the Dutch Preschool Behavior Questionnaire (Dutch PBQ; Goossens, Dekker, Bruinsma, & De Ruyter, 2000) and that aimed at providing a more differentiated view on the EB-dimensions of hyperactivity, physical aggression and opposition. Accordingly, the adapted questionnaire was called the *Hyperactivity-Opposition-Physical-aggression Preschool Assessment (HOPPA; see method section)*.

### **Factorial Invariance for Teacher Questionnaires for Child Externalizing Behavior across Gender and Home Language**

Developing and adapting questionnaires confronts researchers with psychometric questions. For example, for making meaningful comparisons between groups, it is important that, next to factor structure, factorial invariance of a questionnaire is assessed first (Koomen, Verschueren, van Schooten, Jak, & Pianta, 2012; Meredith & Teresi, 2006). Factorial invariance can be examined at subsequent, hierarchical levels (Vandenberg & Lance, 2000). Testing for *configural invariance* reveals whether a questionnaire has the same factor structure for different groups. Investigating *metric invariance* shows whether the strength of the relations between the questionnaire items and the underlying constructs is the same for

different groups. Examining *scalar invariance* establishes whether a construct is measured on a similar scale across groups.

To date, few studies have investigated (simultaneously) the abovementioned levels of factorial invariance for teacher reports of preschool EB across child gender. One notable exception is the study by Spilt and colleagues (2010) who found partial metric invariance for child gender, using the physical aggression items of the Dutch PBQ (Goossens et al., 2000).

Moreover, to date, the population of students at schools becomes more and more linguistically, culturally, and socially diverse, as expressed by the diversity of child home languages in one classroom (Organization for Economic Cooperation and Development, 2009). Therefore, it is important to assess whether teachers interpret children's EB in the same way for children with different home languages. For child ethnicity, which may be seen as a proxy of child home language (Organization for Economic Cooperation and Development, 2009), the scarce studies that do exist yield mixed evidence (Zwirs et al., 2011).

### **The Present Study**

In this study, we sought to verify the a priori three-factor structure (i.e., Hyperactivity, Opposition, Physical aggression) of the HOPPA in a large sample of preschoolers. Moreover, we aimed to investigate whether factorial invariance of the HOPPA held across child gender and across child home language when comparing the majority home language group (i.e., Dutch speaking children) to all other children (i.e., bilingual children and children only speaking a non-Dutch language at home). To ensure sufficient statistical power, the two latter groups were combined in the analyses. Building on Spilt and colleagues (2010), partial metric invariance across gender was expected for Physical aggression, but there were no clear hypotheses for Hyperactivity and Opposition. As there is a dearth of studies investigating

factorial invariance across children with different home languages, factorial invariance was investigated exploratory for different language groups.

## Method

### Participants and Procedure

In the school year 2009-2010, 46 schools were recruited in urban areas of the Flemish region of Belgium (Statistics Belgium, 2012). Parental consent for participation was sought for 3747 preschoolers and obtained for 3610 (96.3%) children from 209 classes (average class size: 17.3). Reasons for non-participation were parental refusal (123 children, 3.3%) and school changes (3 children, 0.1%). For 11 children (0.3%), reasons for not participating were unknown.

The resulting sample consisted of 1776 girls (49.2%) and 1817 boys (50.3%). For 17 children (0.5 %), gender was unknown. Somewhat less than half of the children were in the first preschool group<sup>1</sup> ( $n = 1657$ , 45.9%, age range 2 years and 9 months to 3 years and 8 months), more than half in the second group ( $n = 1934$ , 53.6%, age range 3 years and 9 months to 4 years and 8 months). Preschool group status was unknown for 19 children (0.5%). Most preschoolers ( $n = 2583$ , 71.6%) had Dutch as their home language (i.e., Flanders' official language, spoken in all classrooms), 650 (18%) were bilingual (i.e., Dutch and another language spoken at home), and 195 (5.4%) did not have Dutch as a their home language. For 182 children (5%), home language was unknown. Gender distribution did not vary across home language.

Teachers (100% female) filled out the HOPPA (see Measures) at the end of the school year 2009-2010. Most classes ( $n = 180$ , 86.1%) had one full-time teacher; the other classes ( $n = 29$ , 13.9%) had two part-time teachers.

### Measures

The *Hyperactivity-Opposition-Physical-aggression Preschool Assessment (HOPPA)* was developed to assess preschoolers' Hyperactivity, Opposition, and Physical aggression, as

rated by their teachers. The HOPPA is largely based on the Externalizing scale of the Dutch Preschool Behavior Questionnaire (Dutch PBQ, Goossens et al., 2000), which was, in turn, adapted from the original PBQ (Behar & Stringfield, 1974). The Externalizing scale of the Dutch PBQ (14 items) measures several indicators of child EB, using a 4-point Likert scale ranging from 1 (*Absolutely not characteristic*) to 4 (*Very characteristic*). High internal consistency, interrater agreement, and test-retest stability, as well as concurrent and predictive validity have been shown (Goossens et al., 2000; Goossens et al., 2002). Moreover, the scale has been shown to discriminate between clinical and community samples (Goossens et al., 2000). Spilt and colleagues (2010) distinguished two subscales of 4 items each in this 14-item Externalizing scale, which were called Physical aggression and Non-aggressive antisocial behavior. They found support for the convergent and discriminant validity of these two subscales (Spilt et al., 2010; Spilt, Koomen, Stoel, Thijs, & van der Leij, 2011).

Complementing these authors' work and following the need for targeted screening (e.g., Willoughby et al., 2001) we aimed to distinguish other dimensions of externalizing behavior in the 14 items of the Externalizing scale. For the development of the HOPPA, we chose to retain the 4-item Physical aggression subscale, as the evidence base for this externalizing behavior dimension as a predictor for (later) maladjustment is large (e.g., Côté, Vaillancourt, LeBlanc, Nagin, & Tremblay, 2006; NICHD Early Child Care Research Network, 2004). Next, the Hyperactivity subscale was composed of four other items of the Externalizing scale referring to overactive and impulsive behavior (e.g., 'an overactive child'). Finally, the Opposition subscale was composed of two items from the Externalizing scale (i.e., 'disobedient', 'irritable'), complemented with three items: 'argues a lot', 'stubborn', 'rebellious child'. These three items were added by Spilt and Koomen (J. Spilt, personal communication, December 5, 2009). The first two of those three items were derived from the Teacher Report Form for preschoolers (Achenbach & Rescorla, 2000). The

latter item refers to a commonly used expression for opposition in Dutch. We did not add the four remaining items of the 14-items Externalizing scale referring to more covert and non-aggressive forms of antisocial behavior (e.g., inconsiderate, sneaky) to the HOPPA, because these remaining items could not be adequately assigned to the three dimensions that we aimed to measure.

### **Analytic Plan**

Confirmatory factor analyses (CFA) and multiple group analyses (MGA) were used to examine the factor structure and factorial invariance of the HOPPA using Mplus version 6.1 (Muthén & Muthén, 1998-2012). As the assumption of multivariate normality was not met for most items, robust maximum likelihood estimation was used. Because of substantial intra class correlations (ranging from .09 to .18) and design effects (ranging from 2.43 to 3.86), we controlled for clustering of children in classes using the COMPLEX-function (Muthén, 1994). Model fit was evaluated by the Comparative Fit Index (CFI > .90 for an acceptable and > .95 for a good fit), and the Root Mean Square Error of Approximation (RMSEA < .08 for an acceptable and < .06 for a good fit). The robust chi-square statistic (robust  $\chi^2$ ), which controls for non-normality in the data, is also reported, but this index tends to be biased in large samples (Vandenberg & Lance, 2000).

First, the a priori factor structure (i.e., three correlated factors Hyperactivity, Opposition, and Physical aggression) was tested in two steps. Before doing so, the total sample ( $N = 3610$ ) was randomly divided in Sample 1 ( $n = 1805$ ) and Sample 2 ( $n = 1805$ ). CFA was performed to test the three-factor model in Sample 1. If the model obtained good fit, it was cross-validated in Sample 2.

Second, factorial invariance of the final model across gender and home language was examined (Vandenberg & Lance, 2000). For model identification and factor scaling purposes, the factor loading of the first indicator of each factor (i.e., the reference indicator) was



constrained to one and the intercept of these indicators was constrained to be equal across groups, while factor means and factor variances were freely estimated. The following models were compared: (a) a model in which factor loadings and intercepts of the non-reference items were allowed to differ by group, (Model 0); (b) a model in which the factor loadings of all items were constrained to be equal, but the intercepts of non-reference items were allowed to differ by group (Model 1); and (c) a model in which all factor loadings and intercepts were constrained to be equal (Model 2; Meredith & Teresi, 2006). If Model 0 showed an acceptable fit, *configural invariance* was attained. If Model 1 did not fit the data significantly worse than Model 0, then *metric invariance* was obtained. Additionally, if Model 2 did not fit the data significantly worse than Model 1, *scalar invariance* was supported. Invariance was derived if  $|\Delta CFI| < .02$  (Cheung & Rensvold, 2002), supplemented by  $|\Delta RMSEA| < .015$  (Chen, 2007), when comparing these models.

## Results

### Validating the Three-Factor Model

Model fit of the a priori model with three correlated factors (Hyperactivity, Opposition, Physical aggression) was acceptable in Sample 1 (robust  $\chi^2(62) = 509.436$ , CFI = .921, RMSEA = .063) and in Sample 2 (robust  $\chi^2(62) = 517.467$ , CFI = .916, RMSEA = .064).

In addition, we verified whether a one-factor model (consisting of an overall EB-factor) fitted significantly better than the supposed/hypothesized three-factor model in the total sample. In comparison to the three-factor model (robust  $\chi^2(62) = 824.221$ , CFI = .923, RMSEA = .058), the one-factor model yielded a significantly worse fit ( $|\Delta CFI| = .114$ ,  $|\Delta RMSEA| = .031$ ). Table 1 represents the standardized factor loadings and factor correlations of the three-factor model in the total sample.

### Factorial Invariance across Gender and Home Language

Configural invariance for gender was established as the measurement model yielded an acceptable fit for both boys and girls (see Table 2). Multiple group analysis indicated that both metric ( $|\Delta CFI| = .008$ ,  $|\Delta RMSEA| = .001$ ) and scalar invariance ( $|\Delta CFI| = .013$ ,  $|\Delta RMSEA| = .002$ ) were obtained. Standardized latent mean differences between girls and boys were .458 ( $p < .001$ ) for Hyperactivity, .237 ( $p < .001$ ) for Opposition, and .494 ( $p < .001$ ) for Physical aggression, implying that girls scored significantly lower than boys on these three dimensions.

For home language, configural invariance was attained as the measurement model yielded an acceptable fit in different home language groups (see Table 2). Moreover, multiple group analyses supported both metric ( $|\Delta CFI| = .000$ ,  $|\Delta RMSEA| = .002$ ) and scalar ( $|\Delta CFI| = .006$ ,  $|\Delta RMSEA| = .000$ ) invariance. Standardized latent mean differences between Dutch children and bilingual children or children only speaking a non-Dutch language at home were .107 ( $p = .058$ ) for Hyperactivity, .101 ( $p = .089$ ) for Opposition, and .234 for Physical aggression ( $p < .001$ ), implying that the majority home language group in comparison to all other children scored significantly lower on Physical aggression, but that there were only marginal significant differences between both groups on Opposition and Hyperactivity.

### **Cronbach's Alphas and Observed Mean Scores**

To assess the internal consistency of the HOPPA, Cronbach's alpha's were calculated. In the total sample, Cronbach's  $\alpha$  was .81 for Hyperactivity, .83 for Opposition, and .88 for Physical aggression. The means and standard deviations for Hyperactivity, Physical aggression, and Opposition across different groups can be found in Table 3. Boys on average score significantly higher than girls on Hyperactivity, Physical aggression, and Opposition. The same holds for bilingual children or children only speaking a non-Dutch language at home, on the one hand, in comparison to Dutch children, on the other.

### **Discussion**

The present study aimed to evaluate the factor structure and factorial invariance of a brief, user-friendly, teacher screening questionnaire (the *Hyperactivity-Opposition-Physical-aggression Preschool Assessment questionnaire, HOPPA*) taking a multidimensional approach to preschoolers' EB.

First, the findings supported the distinctiveness and internal consistency of Hyperactivity, Opposition, and Physical aggression. Overall, the three-factor model showed an acceptable fit and a better fit than a one-factor model in two randomly selected samples. Therefore, these findings add to the evidence for the differentiation of EB at a young age (as reported by the teacher; Weis et al., 2005). As these three types of behavior might have a different developmental course and predict differential outcomes (Harvey et al., 2009; Nagin & Tremblay, 1999), it is important that they can be distinctively assessed as early in life as possible. Moreover, the brief, user-friendly nature of the HOPPA may add to easy teacher screening of three relevant externalizing behavior dimensions for all children in the classroom. This, in turn, may add to targeted intervention approaches for these children with high levels on one or more of these externalizing behavior dimensions (Tremblay, 2010).

Second, factorial invariance across gender and child home language was examined for the HOPPA. For both home language and gender, configural, metric, and scalar invariance of the questionnaire was found. This means (a) that the same multi-dimensional structure applies for girls versus boys and for the majority group (i.e., Dutch speaking children) versus all other children (i.e., bilingual children and children only speaking a non-Dutch language at home), (b) that the HOPPA-items are equally good indicators of the underlying EB-dimensions for boys versus girls and the majority home language group versus all other children (i.e., an equal change in the latent construct corresponds to an equal change on the HOPPA-items for both groups), and (c) that a particular score on the HOPPA means exactly the same for boys versus girls and for children of the majority home language group versus all other children

(i.e., scale score differences only reflect differences in underlying latent constructs). In other words, there is no evidence that behaviors (i.e., items) in these different groups of children relate differently to the underlying EB-constructs, nor that teachers interpret children's EB-dimensions differently according to the group to which these children belong.

Notably, finding scalar invariance for gender and for home language is important, as it allows for fair comparisons between the latent means across groups (Cheung & Rensvold, 2002). As such, finding scalar invariance is important both for basic research and for (school psychological) practice (Meredith & Teresi, 2006). For research purposes, for example, finding scalar invariance implies that the results of variance analyses can be validly interpreted when comparing the means of boys versus girls and children of different home language on Hyperactivity, Opposition, and Physical aggression. No bias in the results occurs because of measurement non-invariance across gender or child home language. For (school psychological) practice, finding scalar invariance for the HOPPA implies that this screening instrument allows for fair comparisons across groups. In other words, differences in HOPPA scale scores across groups are not due to biased items but are reflective of true differences in the underlying EB-dimensions.

Although our results for factorial invariance for the HOPPA across gender and home language are promising, only a few studies to date have investigated factorial invariance for questionnaires assessing externalizing behavior dimensions, yielding mixed results. Our study, for instance, provided support for full scalar invariance across gender, while Spilt and colleagues (2010) only found partial metric invariance across gender for Physical aggression (i.e. the scale on which the Physical aggression dimension of the HOPPA is based). One possible explanation may be cultural differences in (teacher perceptions of) gender-specific physical aggression between Belgium (our study) and the Netherlands (study by Spilt et al., 2010). Cultural differences may lead to other stereotypes concerning gender-specific behavior

(e.g., Spilt et al., 2010), explaining these different results. Another possible explanation may be the differences in age of the samples. In our study, the age range of the children was between 2 years and 9 months and 4 years and 8 months, whereas for Spilt et al. (2010)'s study the age range of the two used samples was between 4 years and 10 months and 6 years and 4 months. Stereotypes for gender-specific behavior that teachers hold for Physical aggression may become more stringent as preschoolers grow older. Following these mixed findings, researchers are advised to always assess measurement invariance across gender and home language in their own data set before interpreting gender differences.

In sum, for researchers and practitioners, our study draws attention to the heterogeneity of preschoolers' EB and to the need to use a measure of different EB-dimensions to accurately identify children most at risk for certain outcomes (Nagin & Tremblay, 1999). Some authors suggest that physical aggression in preschool is the most worrisome EB-dimension (Spilt et al., 2010) and that early intervention should focus, first of all, on improving this dimension (Joussemet et al., 2008). However, assessing different EB-dimensions might help us tailor interventions to the specific needs of children, and to develop interventions that prevent a wide variety of outcomes (Tremblay, 2010).

Nevertheless, some limitations of our study should be considered. First, building on the study of Spilt and colleagues (2010) and Spilt and colleagues (2011) and following the call for targeted screening (e.g., Willoughby et al., 2001), our study focused on distinguishing Hyperactivity, Physical aggression and Opposition, which target salient and frequently displayed externalizing behavior dimensions. In previous studies, these externalizing dimensions have been found to be risk factors for maladaptive outcomes later in life (such as delinquency and child diagnoses; Harvey et al., 2009; Nagin & Tremblay, 1999). However, other EB-dimensions (e.g., relational aggression, non-aggressive antisocial behavior) have been found to be predictors of (other forms of) maladjustment too (Crick et al., 2006; Spilt et

al., 2010). Future researchers may try to assess these other dimensions by adding key indicators of these concepts to the HOPPA, while preserving the conciseness of the questionnaire.

Second, we only used teacher questionnaires. Although teachers have been shown to provide reliable and valid reports on EB (Konold & Pianta, 2007), future research could take a multitrait-multimethod matrix approach (Campbell & Fiske, 1959) to confirm the convergent and divergent validity of the different HOPPA-subscales. For Physical aggression, a previous study of Spilt and colleagues (2011) already confirmed the convergent and divergent validity for Physical aggression. Future research should do the same for Hyperactivity and Opposition.

Third, in this study, all teachers were female. Although in most preschool classes (in Flanders) teachers are female, the gender of the teacher might have had an influence on the reports of preschool EB. This should be considered when using the HOPPA as a screening instrument. On a related note, the HOPPA was developed for research and screening purposes, not for diagnostic use.

Fourth, for reasons of statistical power, bilingual children and children only speaking a non-Dutch language at home were combined in the factorial invariance analyses. Children in both groups are likely to have at least one parent with a non-Belgian ethnicity. Variations in socio-cultural and socio-linguistic beliefs may influence these children's interactions with their teacher (e.g., Oades-Sese & Li, 2011). Moreover, speaking another language at home may lead to confusion and misunderstanding. As language serves as a bonding agent between children and their attachment figures, such as their teachers (Oades-Sese & Li, 2011), speaking a different language at home may negatively affect teacher-child relationship quality. Despite these common characteristics that may affect teacher perceptions of child behavior (e.g., Doumen et al., 2008), it is recommended that future research on factorial

invariance treats bilingual children and children who do not speak the official language of the school as two separate groups.

Fifth, as the main goal of the HOPPA is to accurately identify preschoolers at risk for maladjustment, the predictive value of the Hyperactivity, Opposition, and Physical aggression scales for future maladjustment should be assessed. Although previous research has shown that hyperactivity, opposition, and physical aggression have at least partly a different developmental course and predict different maladjustment outcomes (e.g., Harvey et al., 2009; Nagin & Tremblay, 2009), future research should assess whether the same holds when these dimensions are assessed with the HOPPA.

In sum, the current study addressed the call of scholars for a brief, user-friendly screening questionnaire focused on assessing different EB-dimensions rather than using a more generic broadband scale of EB. This study is the first to indicate the usefulness of the HOPPA as a brief screening measure to obtain information from teachers about preschoolers' hyperactivity, opposition, and physical aggression. Teachers assigned the same meaning to these constructs for boys and girls and for children with different home languages.

### References

- Achenbach, T. M., & Rescorla, L. A. (2000). *Manual for the ASEBA Preschool Forms & Profiles*. Burlington, VT: University of Vermont.
- Behar, L., & Stringfield, S. (1974). A behavior rating scale for the preschool child. *Developmental Psychology, 10*, 601-610. doi: 10.1037/h0037058
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*, 81-105. doi: 10.1037/h0046016
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 14*, 464-504.
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*, 233-255. doi: 10.1207/S15328007SEM0902\_5
- Côté, S., Vaillancourt, T., LeBlanc, J., Nagin, D., & Tremblay, R. (2006). The development of physical aggression from toddlerhood to pre-adolescence: A nation-wide longitudinal study of Canadian children. *Journal of Abnormal Child Psychology, 34*, 68-82. doi: 10.1007/s10802-005-9001-z
- Crick, N. R., Ostrov, J. M., Burr, J. E., Cullerton-Sen, C., Jansen-Yeh, E., & Ralston, P. (2006). A longitudinal study of relational and physical aggression in preschool. *Applied Developmental Psychology, 27*, 254-268. doi: 10.1016/j.appdev.2006.02.006
- Dever, B.V., & Kamphaus, R.W. (2013). Behavioral, social, and emotional assessment of children. In K. F. Geisinger, B. A. Bracken, J. F. Carlson, J. C. Hansen, N. R. Kuncel, S. P. Reise, & M. C. Rodriguez (Eds.), *APA Handbook of Testing and Assessment in Psychology: Testing and Assessment in School Psychology and Education* (Vol 3, pp. 129-148). Washington, DC: American Psychological Association (APA).



- Doumen, S., Verschueren, K., Buyse, E., Germeijs, V., Luyckx, K., & Soenens, B. (2008). Reciprocal relations between teacher-child conflict and aggressive behaviour in kindergarten: A three-wave longitudinal study. *Journal of Clinical Child and Adolescent Psychology*, 37, 588–599. doi: 10.1080/15374410802148079
- Feil, E. G., Walker, H. M., & Severson, H. H. (1995). The early screening project for young children with behavior problems. *Journal of Emotional and Behavioral Disorders*, 3, 194-202. doi: 10.1177/106342669500300401
- Goossens, F. A., Bokhorst, K., Bruinsma, C., & Van Boxtel, H. W. (2002). Judgments of aggressive, withdrawn and prosocial behavior: Perceived control, anger, pity and sympathy in young Dutch children. *Journal of School Psychology*, 40, 309-327. doi: 10.1016/S0022-4405(02)00103-6
- Goossens, F. A., Dekker, P., Bruinsma, C., & De Ruyter, P. A. (2000). *De Gedrags Vragenlijst voor Peuters en Kleuters: Factor structuur, betrouwbaarheid en validiteit*. [Behavior questionnaire for 2- to 6-year-olds: Factor structure, reliability, and validity]. Amsterdam, the Netherlands: Vrije Universiteit.
- Harvey, E. A., Youngwirth, S. D., Thakar, D. A., & Errazuriz, P. A. (2009). Predicting Attention Deficit/Hyperactivity Disorder and Oppositional Defiant Disorder from preschool diagnostic assessments. *Journal of Consulting and Clinical Psychology*, 77, 349-354. doi: 10.1037/a0014638
- Joussemet, M., Vitaro, F., Barker, E. D., Côté, S., Nagin, D. S., Zoccolillo, M., & Tremblay, R. E. (2008). Controlling parenting and physical aggression during elementary school. *Child Development*, 79, 411-425. doi: 10.1111/j.1467-8624.2007.01133.x
- Koomen, H. M. Y., Verschueren, K., van Schooten, E., Jak, S., & Pianta, R. C. (2012). Validating the Student-Teacher Relationship Scale: Testing factor structure and

- measurement invariance across child gender and age in a Dutch sample. *Journal of School Psychology, 50*, 215-234. doi: 10.1016/j.jsp.2011.09.001
- Konold, T. R., & Pianta, R. C. (2007). The influence of informants on the ratings of children's behavioral functioning: A latent variable approach. *Journal of Psychoeducational Assessment, 25*, 222-236. doi: 10.1177/0734282906297784
- Loeber, R., & Farrington, D. P. (2000). Young children who commit crime: Epidemiology, developmental origins, risk factors, early interventions, and policy implications. *Development and Psychopathology, 12*, 737-762. doi: 10.1017/S0954579400004107
- Matthijs, W., & Lochman, J. E. (2010). *Oppositional defiant disorder and conduct disorder in childhood*. UK: John Wiley & Sons.
- Meredith, W. M., & Teresi, J. A. (2006). An essay on measurement and factorial invariance. *Medical Care, 44*, S69-S77. doi: 10.1097/01.mlr.0000245438.73837.89
- Moreland, A. D., & Dumas, J. E. (2008). Categorical and dimensional approaches to the measurement of disruptive behavior in the preschool years: A meta-analysis. *Clinical Psychology Review, 28*, 1059-1070. doi: 10.1016/j.cpr.2008.03.001
- Muthén, B. (1994). Multilevel covariance structure analysis. *Sociological Methods & Research, 22*, 376-398. doi: 10.1177/0049124194022003006
- Muthén, L. K., & Muthén, B. O. (1998-2012). *Mplus user's guide (6th ed.)*. Los Angeles: Author.
- Nagin, D., & Tremblay, R. E. (1999). Trajectories of boys' physical aggression, opposition, and hyperactivity on the path to physically violent and nonviolent juvenile delinquency. *Child Development, 70*, 1181-1196. doi: 10.1111/1467-8624.00086
- NICHD Early Child Care Research Network (2004). Trajectories of physical aggression from toddlerhood to middle childhood. *Monographs of the Society for Research in Child Development, 96*(4, Serial No. 278).

- Oades-Sese, G. V., & Li, Y. (2011). Attachment relationships as predictors of language skills for at-risk bilingual preschool children. *Psychology in the Schools, 48*, 707-722. doi: 10.1002/pits.20583
- Organization for Economic Cooperation and Development (OECD) (2009). *Programme for International Student Assessment (PISA) results: Overcoming social background: Equity in learning opportunities and outcomes*. Retrieved from <http://www.oecd.org/pisa/pisaproducts/48852584.pdf>
- Spilt, J. L., Koomen, H. M. Y., Thijs, J. T., Stoel, R. D., & van der Leij, A. (2010). Teachers' assessment of antisocial behavior in kindergarten. *Journal of Psychoeducational Assessment, 28*, 129-138. doi: 10.1177/0734282909340236
- Spilt, J. L., Koomen, H. M. Y., Stoel, R. D., Thijs, J. T., & van der Leij, A. (2011). Teachers' assessment of physical aggression with the Preschool Behavior Questionnaire: A Multitrait-Multimethod Evaluation of convergent and discriminant validity. *Journal of Psychoeducational Assessment, 29*, 407-417. doi: 10.1177/0734282910388881
- Statistics Belgium (2012). *Typologie van gemeenten volgens twee verschillende bronnen: Eurostat en OESO [Typology of areas according to two sources: Eurostat and OECD]*. Belgium: Statistics Belgium.
- Tremblay, R. E. (2010). Developmental origins of disruptive behaviour problems: the “original sin” hypothesis, epigenetics and their consequences for prevention. *Journal of Child Psychology and Psychiatry, 51*, 341-367. doi: 10.1111/j.1469-7610.2010.02211.x
- Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*, 4-70. doi: 10.1177/109442810031002

- Weis, R., Lovejoy, M. C., & Lundahl, B. W. (2005). Factor structure and discriminative validity of the Eyberg Child Behavior Inventory with young children. *Journal of Psychopathology and Behavioral Assessment*, 27, 269-278. doi: 10.1007/s10862-005-2407-7
- Willoughby, M., Kupersmidt, J., & Bryant, D. (2001). Overt and covert dimensions of antisocial behavior in early childhood. *Journal of Abnormal Child Psychology*, 29, 177-187. doi: 10.1023/A:1010377329840
- Zwirs, B., Burger, H., Schulpen, T., Vermulst, A. A., HiraSing, R. A., & Buitelaar, J. (2011). Teacher ratings of children's behavior problems and functional impairment across gender and ethnicity: Construct equivalence of the Strengths and Difficulties Questionnaire. *Journal of Cross-Cultural Psychology*, 42, 466-481. doi: 10.1177/0022022110362752

**Footnotes**

<sup>1</sup> In Belgium, preschool education is available for children aged 2.5 to 6 years. Children between the ages of 2.5 and 3 are mostly taught in a separate class (a sort of pre-preschool class). The first, second and third preschool group respectively consist of children aged between 2 years and 9 months and 3 years and 8 months, between 3 years and 9 months and 4 years and 8 months, and between 4 years and 9 months and 5 years and 8 months. With the term “preschool group status” we refer to differences in the preschool group.

Table 1

*Items of the HOPPA: Standardized Factor loadings and Factor Correlations of the Three-Factor Model (13 items) in the Total Sample (N = 3610)*

	Hyperactivity	Physical aggression	Opposition	$R^2$
Restless	.765			.585
Runs about, jumps up and down	.781			.610
An overactive child	.889			.790
Poor concentration	.469			.220
Kicks, hits		.875		.766
Bullies		.795		.632
Fights		.869		.755
Destroys things		.688		.473
Disobedient			.809	.654
Irritable			.608	.370
Stubborn			.642	.412
Argues a lot			.717	.514
A rebellious child			.784	.615
Factor correlations	Hyperactivity	Physical aggression		
Hyperactivity	-			
Physical aggression	.762	-		
Opposition	.767	.762		

Table 2

*Measurement Invariance of the Three-Factor Model for Gender and Home Language*

	df	robust $\chi^2$	CFI	RMSEA
<i>Configural invariance</i>				
Dutch home language	62	569.388	0.928	0.056
Bilingual/Other language	62	340.938	0.919	0.073
Males	62	595.242	0.922	0.069
Females	62	334.647	0.928	0.050
<i>Metric and scalar invariance</i>				
Home language variant	124	935.360	0.925	0.062
Home language metric	134	951.247	0.925	0.060
Home language scalar	144	1022.716	0.919	0.060
Gender variant	124	906.058	0.924	0.059
Gender metric invariant	134	997.914	0.916	0.060
Gender scalar invariant	144	1140.192	0.903	0.062

*Note.* df = degrees of freedom; robust  $\chi^2$  = robust chi square; CFI = Comparative Fit Index; RMSEA = Root Mean Square Error of Approximation.

Table 3

*Means and Standard Deviations of Hyperactivity, Physical Aggression, and Opposition across Girls, Boys, Dutch, and Bilingual or Non-Dutch Children*

Variable	Gender					Home language				
	Girls ( <i>n</i> = 1752)		Boys ( <i>n</i> = 1758)		<i>t</i> test	Dutch ( <i>n</i> = 2583)		Bilingual or Non-Dutch ( <i>n</i> = 845)		<i>t</i> test
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
Hyper	1.33	0.51	1.65	0.72	-15.43***	1.47	0.63	1.57	0.71	-3.87***
PA	1.16	0.35	1.45	0.65	-16.63***	1.27	0.50	1.42	0.67	-6.03***
Opposition	1.34	0.49	1.45	0.59	-6.31***	1.39	0.54	1.43	0.57	-2.07*

*Note.* The variations in sample size are due to missings. Hyper = Hyperactivity. PA = Physical aggression.

\**p*<.05. \*\**p*<.01. \*\*\**p*<.001.